Superquantile/CVaR Risk Measures: Second-Order Theory[1]

R. Tyrrell Rockafellar   *Johannes O. Royset*

Department of Mathematics   Operations Research Department
University of Washington   Naval Postgraduate School
rtr@uw.edu   joroyset@nps.edu

Abstract.   Superquantiles, which refer to conditional value-at-risk (CVaR) in the same way that quantiles refer to value-at-risk (VaR), have many advantages in the modeling of risk in finance and engineering. However, some applications may benefit from a further step, from superquantiles to second-order superquantiles. Measures of risk based on second-order superquantiles have recently been explored in some settings, but key parts of the theory have been lacking: descriptions of the associated risk envelopes and risk identifiers. Those missing ingredients are supplied in this paper, and moreover not just for second-order superquantiles, but also for a much broader class of mixed superquantile measures of risk. Such dualizing expressions facilitate the development of dual methods for mixed and second-order superquantile risk minimization as well as superquantile regression, a proposed second-order version of quantile regression.

Keywords: superquantiles, conditional value-at-risk, second-order superquantiles, mixed superquantiles, spectral measures of risk, risk envelopes, risk identifiers, duality of risk measures, superquantile regression

Date: November 5, 2015

1 Introduction

The second-order version of conditional value-at-risk that we introduced in [21], with further explanations in [20, 23], corresponds to a sort of smoothing of the cumulative distribution function of a random variable but has other key interpretations as well. Motivated by specific applications in risk-averse optimization and regression for physical systems, we develop it further here with particular attention

---

to duality. The term "superquantile" as an alternative to "conditional value-at-risk" [19] is employed for this broad purpose, beyond the usual domain of finance.

To understand the second-order ideas with which we will be occupied in this paper, some background in the first-order ideas is needed, and we begin briefly with that. The conditional value-at-risk $\text{CVaR}_\alpha(X)$ of a random variable X oriented to "loss" or "cost," at a probability level $\alpha \in [0, 1)$, is the expected value of the α-upper tail distribution of X as defined in [24, 25]. When the cumulative distribution function F_X for X is continuous at $\text{VaR}_\alpha(X)$, the value-at-risk of X at level α, this tail distribution is simply the conditional distribution for X with respect to the interval $[\text{VaR}_\alpha, \infty)$, but otherwise it requires taking into account an atom of probability at $\text{VaR}_\alpha(X)$. This distinction makes CVaR different from other notions introduced around the same time, such as "tail-VaR" [2], which includes the entire probability atom, and "mean shortfall" [12], which omits it (although the similar term "expected shortfall" has been ambiguous in this respect). Conditional value-at-risk can also be expressed by the formula

$$\text{CVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_\beta(X) d\beta$$

of [1], which was adopted by Föllmer and Schied as the definition of "average" value-at-risk [7].[2]

These other concepts were originally articulated for random variables oriented toward gain, but the loss orientation we follow here has the advantage of making the value-at-risk $\text{VaR}_\alpha(X)$ coincide with the α-*quantile* $q_X(\alpha)$ familiar in statistics:

$$\text{VaR}_\alpha(X) = q_X(\alpha) = \min\{x \in I\!\!R | F_X(x) \geq \alpha\}.$$

This uniting of VaR with quantiles has further suggested a way of exiting from finance-driven terminology about risk for the sake of applications outside of finance, namely by speaking of the conditional value-at-risk $\text{CVaR}_\alpha(X)$ as the α-*superquantile* of X in the parallel notation $\bar{q}_X(\alpha)$. Then the integral formula for $\text{CVaR}_\alpha(X)$ becomes

$$\bar{q}_X(\alpha) = \frac{1}{1-\alpha} \int_\alpha^1 q_X(\beta) d\beta. \tag{1}$$

With this shift we have a platform for displaying the *second-order superquantiles* of [21], to be denoted by $\bar{\bar{q}}_X(\alpha)$; they are defined by

$$\bar{\bar{q}}_X(\alpha) = \frac{1}{1-\alpha} \int_\alpha^1 \bar{q}_X(\beta) d\beta. \tag{2}$$

There is more to the second-order superquantile than just the analogy between (1) and (2), though, as has been laid out in [21].

Especially of interest is a formula derived in [21] that extends to superquantiles and second-order superquantiles the basic connection between VaR and CVaR discovered in [24, 25]. That earlier formula asserts, in quantile/superquantile notation, that

$$\bar{q}_X(\alpha) = \min_{c \in \mathbf{R}}\{c + \mathcal{V}_\alpha(X - c)\}, \qquad q_X(\alpha) = \operatorname*{argmin}_{c \in \mathbf{R}}\{c + \mathcal{V}_\alpha(X - c)\}, \tag{3}$$

---

[2]They preferred "average" because "conditional" could have differing usages. This issue also adds motivation to our passage to "superquantiles."

in terms of the "regret" functional

$$\mathcal{V}_\alpha(X) = \frac{1}{1-\alpha} E[\max\{0, X\}] = \frac{1}{1-\alpha} \int_0^1 \max\{0, q_X(\beta)\} d\beta$$

(with the argmin being an interval which, if not a singleton, has the quantile in question as its left endpoint). The second-order extension asserts that

$$\bar{\bar{q}}_X(\alpha) = \min_{c \in \mathbf{R}}\{c + \bar{\mathcal{V}}_\alpha(X - c)\}, \qquad \bar{q}_X(\alpha) = \operatorname*{argmin}_{c \in \mathbf{R}}\{c + \bar{\mathcal{V}}_\alpha(X - c)\}, \tag{4}$$

where

$$\bar{\mathcal{V}}_\alpha(X) = \frac{1}{1-\alpha} \int_0^1 \max\{0, \bar{q}_X(\beta)\} d\beta. \tag{5}$$

Achieving such a formula had been one of our main goals in pursuing second-order superquantilies, because it is deeply tied to generalized regression. The joint formula (3) is central to *quantile* regression, a well known alternative to ordinary least-squares regression, so the joint formula (4) indicates a possible elevation to *superquantile* regression. The double formula (4) was developed in [21] through a technique in which the superquantiles of X could be interpreted as the quantiles of a "super" random variable $\bar{X}$ associated with X.

Our primary aim in this paper is to fill in missing parts of the second-order theory concerned with duality. An important ingredient of duality for any coherent measure of risk, including superquantile/CVaR, is an expression of the risk as a worst-case expectation over an associated class of probability measures. This requires identifying the "risk envelope" that characterizes that class. The risk envelope for the risk measure given by second-order superquantiles has not yet been fully understood, but we will pin it down here.

This pushes us naturally into wider terrain in observing that the integral formula for the second-order superquantile casts it as a special "spectral" measure of risk of X in the sense of Acerbi [1]. Spectral measures of risk, which have also been studied from various angles under the heading of mixed superquantile/CVaR measures of risk [28, 26], are known to be fundamental for characterizing coherent measures of risk that are law-invariant [11, 7, 10, 17, 15, 32]. Some properties of risk envelopes of mixed superquantile measures of risk are known for the setting with random variables defined on a nonatomic probability space; see [9, Sections 4.5 and 4.6] and [32]. We provide a development for arbitrary probability spaces and give explicit formulae for risk envelopes and identifiers, especially for second-order superquantile risk measures. Moreover, in the broader framework of risk quadrangles [26], we provide a deeper understanding of the risk quadrangles with (first-order) superquantile/CVaR as its statistics and for the first time state explicit expressions for risk envelopes and identifiers in the case of (finite) mixture of quantiles as the statistic.

Although dualization of risk measures can be carried out for a variety of spaces of random variables and paired dual spaces (see for example [31, 4, 9]), we focus here on random variables with finite second moments. This excludes some applications, for example in finance, where only the first moments are finite. Still, in many applications in engineering and generalized regression, the assumption appears tolerable. A compelling reason in our setting is that this restriction guarantees the finiteness of second-order superquantiles. That follows from their expression as an integral of first-order superquantiles and the bounds derived for the latter under such restriction in [23, Proposition 1], namely

$$E[X] \le \bar{q}_X(\alpha) \le \min\left\{E[X] + \frac{\sigma(X)}{\sqrt{1-\alpha}}, \ \sup X\right\}, \tag{6}$$

where $\sigma(X)$ denotes standard deviation and the lower bound is strict for nonconstant X unless $\alpha = 0$. Another plus is that this choice allows for random variables with normal distributions, whereas much of the literature in finance restricts consideration to random variables with essentially bounded range.

In Section 2 we lay the foundation for working in this framework and the measures of risk that fit into it. We proceed in Section 3 with the central results of duality concerning risk envelopes and the risk identifiers they associate with random variables. Section 4 then applies the results to optimization and generalized regression. An appendix collects some of the technical details that are needed along the way.

2 Risk Measure Framework

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we let

$$\mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) := \{X : \Omega \to I\!R \mid X \ \ \mathcal{F}\text{-measurable}, \ E[X^2] < \infty\}$$

be the space of random variables with finite second moment, where we write integration with respect to $\mathbb{P}$ using the standard notation $E[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega)$. We equip $\mathcal{L}^2$ with the standard norm $\|X\|_2 := (E[X^2])^{1/2}$. As explained in the introduction, the choice of $\mathcal{L}^2$ ensures, through (6), the finiteness of the second-order superquantiles $\bar{\bar{q}}_X(\alpha)$ we are especially focused on.

In the following, we deal with classes of measures of risk defined on $\mathcal{L}^2$. Regularity [26, 22] provides fundamental properties for such risk measures. We recall that a measure of risk $\mathcal{R} : \mathcal{L}^2 \to (-\infty, \infty]$ is *regular* if it satisfies the following axioms:

$\mathcal{R}(X) = c$ for constant random variables $X \equiv c$,

$\mathcal{R}((1-\tau)X + \tau X') \leq (1-\tau)\mathcal{R}(X) + \tau \mathcal{R}(X')$ for all $X, X' \in \mathcal{L}^2$ and $\tau \in (0,1)$ (convexity),

$\{X \in \mathcal{L}^2 \mid \mathcal{R}(X) \leq c\}$ is closed for all $c \in I\!R$ (closedness),

$\mathcal{R}(X) > E[X]$ for nonconstant $X \in \mathcal{L}^2$ (averseness),

which have as a consequence that $\mathcal{R}(X + c) = \mathcal{R}(X) + c$ for all $c \in I\!R$. In fact we will only be working here with risk measures that in addition are both *positively homogeneous*,

$$\mathcal{R}(\tau X) = \tau \mathcal{R}(X) \text{ for } \tau > 0, \ X \in \mathcal{L}^2,$$

and *monotonic*,

$$\mathcal{R}(X) \leq \mathcal{R}(Y) \text{ whenever } X(\omega) \leq Y(\omega) \text{ for a.e. } \omega \in \Omega.$$

In particular $\mathcal{R}$ is then a *coherent* measure of risk in the sense of [2]. Duality in this case is expressed by the following correspondence between risk measures $\mathcal{R}$ and sets $\mathcal{Q}$ called their *risk envelopes*.[3]

2.1 Proposition (risk envelope duality). *For a regular measure of risk $\mathcal{R}$ on $\mathcal{L}^2$ that is positively homogeneous and monotone, the relations*

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} E[XQ] \text{ for } X \in \mathcal{L}^2, \quad \mathcal{Q} = \{Q \in \mathcal{L}^2 \mid E[XQ] \leq \mathcal{R}(X) \text{ for all } X \in \mathcal{L}^2\},$$

---

[3]The term "risk envelope" was introduced in 2002 in [27].

give a one-to-one correspondence between the regular measures of risk $\mathcal{R}$ on $\mathcal{L}^2$ that are positively homogeneous and monotonic and the nonempty closed convex subsets $\mathcal{Q}$ of $\mathcal{L}^2$ that consist of elements $Q \geq 0$ with $E[Q] = 1$ and are such that each nonzero $X \in \mathcal{L}^2$ has $E[XQ] > 0$ for at least one $Q \in \mathcal{Q}$.

This fact, a specialization of the general support function correspondence in convex analysis, is a variant from [26] of known results characterizing other classes of risk measures, starting with [2]. Important along with the risk envelope $\mathcal{Q}$ associated with $\mathcal{R}$ are the sets

$$\mathcal{Q}^X = \operatorname*{argmax}_{Q \in \mathcal{Q}} E[XQ] \text{ for } X \in \mathcal{L}^2, \tag{7}$$

which are called the *risk identifiers* for the individual random variables X.[4]

The measures of risk at the center of our attention are the *first-order* superquantile measures $\mathcal{R}_\alpha$ and the *second-order* superquantile measures $\bar{\mathcal{R}}_\alpha$ given by

$$\mathcal{R}_\alpha(X) = \bar{q}_X(\alpha) \quad \text{and} \quad \bar{\mathcal{R}}_\alpha(X) = \bar{\bar{q}}_X(\alpha) \quad \text{for } \alpha \in [0,1) \tag{8}$$

in accordance with the expressions (1) for $\bar{q}_X(\alpha)$ and (2) for $\bar{\bar{q}}_X(\alpha)$ in Section 1. The properties of $\mathcal{R}_\alpha$ that make it regular, positively homogeneous, and monotonic have been known for some time, and those properties are obviously inherited by $\bar{\mathcal{R}}_\alpha$ through the expression of $\bar{\bar{q}}_X(\alpha)$ as an integral in (2). In both cases, therefore, we are dealing with measure of risk covered by the preceding theorem. For $\mathcal{R}_\alpha$, the risk envelope is known to be

$$\mathcal{Q}_\alpha := \{ Q \in \mathcal{L}^2 \mid 0 \leq Q(\omega) \leq 1/(1-\alpha) \text{ a.e. } \omega \in \Omega, \; E[Q] = 1 \}, \tag{9}$$

cf. [27, 26]. For $\bar{\mathcal{R}}_\alpha$, the specifics of the corresponding risk envelope $\bar{\mathcal{Q}}_\alpha$ will be determined for the first time in Section 3; some properties are known from [9, Sections 4.5 and 4.6] and [32] under the additional assumption that $(\Omega, \mathcal{F}, \mathbb{P})$ is nonatomic. However, to accomplish this efficiently and gain other new insights at the same time, we will pass through a broader class of risk measures that we call mixed superquantile measures of risk. Such measures are already known, having been identified by Acerbi [1] as corresponding broadly to "spectral" measures in the pattern of dual utility theory. They are a key ingredient in the sup representation that Kusuoka [11] has provided for a class of law-invariant measures of risk, later also studied in [16, 32].

2.2 Definition (mixed superquantile measures of risk). *For a weighting measure λ, namely a probability measure on $([0,1), \mathcal{B}_{[0,1)})$,[5] the associated mixed superquantile measure of risk $\mathcal{R} : \mathcal{L}^2 \to (-\infty, \infty]$ is given by*

$$\mathcal{R}(X) := \int_0^1 \bar{q}_X(\beta) \, d\lambda(\beta). \tag{10}$$

For technical reasons, we exclusively deal in this situation with the completion of $([0,1), \mathcal{B}_{[0,1)}, \lambda)$, which, with a slight abuse of notation, we denote by $([0,1), \bar{\mathcal{B}}_{[0,1)}, \lambda)$.

---

[4]This term was introduced in [28], although the sets in question were handled earlier as being the subdifferentials of convex analysis for the risk measure functionals in question.

[5]For a set S with a topology, let $\mathcal{B}_S$ be its Borel sigma-algebra.

Here, the key aspect is that second-order superquantile risk measures fit this definition because, through (2), we have

$$\bar{\mathcal{R}}_\alpha(X) = \bar{\bar{q}}_X(\alpha) = \int_0^1 \bar{q}_X(\alpha) \, d\bar{\lambda}_\alpha(\beta), \text{ where } \bar{\lambda}_\alpha(S) := \frac{m(S \cap (\alpha, 1))}{1 - \alpha} \text{ for } S \in \mathcal{B}_{[0,1)}.^6 \qquad (11)$$

As another special case, if λ is concentrated on a finite number of points in $[0, 1)$, say $\alpha_1, \alpha_2, ..., \alpha_k$, then simply $\mathcal{R}(X) = \lambda(\alpha_1)\bar{q}_X(\alpha_1) + \cdots + \lambda(\alpha_k)\bar{q}_X(\alpha_k)$. A first-order superquantile risk measure is realized by setting $k = 1$.

Note in general that, since λ is defined on $\mathcal{B}_{[0,1)}$, we exclude the possibility of a weighting measure that places a positive weight at $\alpha = 1$. That case simply yields $\mathcal{R}(X) = \infty$ whenever $\sup X = \infty$, and it is better treated separately.

The basic properties of a mixed superquantile risk measure are described by the following result, where certain parts are immediate from the definition. The result extends previous results in [27, 28], which built in turn on the earlier spectral representation of Acerbi [1], by dealing with a significantly relaxed condition for finiteness and admitting the point $\beta = 0$ explicitly.

2.3 Proposition (mixed superquantile properties). *A mixed superquantile risk measure $\mathcal{R}$ as in (10) is well-defined, monotonic and positively homogeneous. It is regular if $\lambda(\{0\}) < 1$, but lacking averseness if $\lambda(\{0\}) = 1$. Specifically,*

$$\mathcal{R}(X) \geq E[X] \text{ for all } X \in \mathcal{L}^2 \text{ and } \mathcal{R}(X) > E[X] \text{ for nonconstant } X \text{ unless } \lambda(\{0\}) = 1.$$

It is finite on $\mathcal{L}^2$ whenever the weighting measure λ satisfies

$$\int_0^1 \frac{1}{\sqrt{1 - \beta}} \, d\lambda(\beta) < \infty$$

and, regardless of the weighting measure, has $\mathcal{R}(X) < \infty$ whenever $\sup X < \infty$.

It has the alternative expression

$$\mathcal{R}(X) = \int_0^1 q_X(\beta)\varphi(\beta)d\beta, \text{ where } \varphi(\beta) := \int_{0 \leq \alpha < \beta} \frac{1}{1 - \alpha} \, d\lambda(\alpha), \ \beta \in [0, 1].$$

The risk profile function φ is right-continuous and nondecreasing on $[0, 1]$ with $\varphi(0) = 0$ and satisfies $\int_0^1 (1 - \alpha)d\varphi(\alpha) = 1$. Conversely, any φ with these properties arises from a unique weighting measure λ given by $d\lambda(\alpha) = (1 - \alpha)d\varphi(\alpha)$.

The proof of this proposition, similar in some ways to that of previous versions but containing new parts, is provided in the Appendix. Further clarification of properties of mixed superquantile measures of risk has been furnished in [26, Mixing Theorem].

Next on the agenda is applying this general result to the case in (11) that corresponds to second-order superquantile measures of risk.

---

[6] Here, and throughout the paper, m denotes Lebesgue measure.

2.4 Theorem (second-order superquantile properties). *Any second-order superquantile risk measure* $\bar{\mathcal{R}}_\alpha : \mathcal{L}^2 \to \mathbb{R}$, $\alpha \in [0, 1)$, *is regular, monotonic, and positively homogenous, and satisfies for* $X \in \mathcal{L}^2$

$$E[X] \leq \bar{q}_X(\alpha) = \bar{\mathcal{R}}_\alpha(X) \leq \min\left\{ E[X] + \frac{2\sigma(X)}{\sqrt{1-\alpha}}, \ \sup X \right\},$$

with the lower bound holding with strict inequality whenever X *is nonconstant.*

It *has the alternative expressions*

$$\bar{\mathcal{R}}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 q_X(\beta) \log \frac{1-\alpha}{1-\beta} d\beta = \int_0^1 q_X(\beta) \bar{\varphi}_\alpha(\beta) d\beta,$$

with respect to the risk profile function

$$\bar{\varphi}_\alpha(\beta) := \begin{cases} \frac{1}{1-\alpha} \log \frac{1-\alpha}{1-\beta} & \text{if } \alpha \leq \beta < 1 \\ 0 & \text{if } 0 \leq \beta < \alpha. \end{cases}$$

Moreover, $\bar{\varphi}_\alpha$ *is a nondecreasing, finite convex function on* $[0, 1]$ *with right-derivative equal to* $1/(1-\alpha)^2$ *as it starts to grow from 0 at* $\beta = \alpha$.

Proof. As a special case of Proposition 2.3, it follows automatically that $\bar{\mathcal{R}}_\alpha$ is well-defined, regular, monotonic, positively homogeneous, and bounded from below by $E[X]$. From (6),

$$\bar{\mathcal{R}}_\alpha(X) \leq \frac{1}{1-\alpha} \int_\alpha^1 E[X] + \frac{\sigma(X)}{\sqrt{1-\beta}} d\beta = E[X] + \frac{\sigma(X)}{1-\alpha} \int_\alpha^1 \frac{1}{\sqrt{1-\beta}} \, d\lambda(\beta) = E[X] + \frac{2\sigma(X)}{\sqrt{1-\alpha}}.$$

Obviously, $\bar{\mathcal{R}}_\alpha(X) \leq \sup X$ also holds.

The alternative expression follows after a specialization of φ of Proposition 2.3 for the given choice of weighting measure $\lambda = \bar{\lambda}_\alpha$. Specifically,

$$\varphi(\beta) = \int_{0 \leq \gamma < \beta} \frac{1}{1-\gamma} d\bar{\lambda}_\alpha(\gamma) = \bar{\varphi}_\alpha(\beta) = \begin{cases} \int_\alpha^\beta \frac{1}{1-\gamma} \frac{1}{1-\alpha} d\gamma & \text{if } \alpha \leq \beta < 1 \\ 0 & \text{if } 0 \leq \beta \leq \alpha. \end{cases}$$

Since for $0 \leq a \leq b < 1$,

$$\int_a^b \frac{1}{1-\beta} d\beta = \log \frac{1-a}{1-b},$$

we therefore find that the alternative expressions follow.

The assertion about $\bar{\varphi}_\alpha$ being convex is justified by its derivative being zero for $\beta \in (0, \alpha)$ and $1/((1-\alpha)(1-\beta))$ for $\beta \in (\alpha, 1)$, with left- and right-derivatives at $\beta = \alpha$ equal to 0 and $1/(1-\alpha)^2$, respectively. $\qquad\square$

The upper bounds on $\mathcal{R}_\alpha$ (see (6)) and $\bar{\mathcal{R}}_\alpha$ (Theorem 2.4), the latter given here for the first time, are remarkably similar. They show that, although second-order superquantile risks are larger than first-order risks, the difference is at most $\sigma(X)/\sqrt{1-\alpha}$.

3   Dualization Through Risk Envelopes

We now turn to determining the dual expressions for mixed and second-order superquantile risk measures in terms of the risk envelopes described in general in Proposition 2.1. The risk envelope $\mathcal{Q}_\alpha$ that corresponds to the first-order superquantile measure $\mathcal{R}_\alpha$ in (8) has already been indicated in (9).

Another case where the risk envelope is already known is that of a mixed superquantile measure $\mathcal{R}$ associated with a weighting measure λ that is concentrated in finitely many points. Namely, if $\mathcal{R} = \lambda_1 \mathcal{R}_{\alpha_1} + \cdots + \lambda_k \mathcal{R}_{\alpha_k}$, the corresponding risk envelope is $\mathcal{Q} = \lambda_1 \mathcal{Q}_{\alpha_1} + \cdots + \lambda_k \mathcal{Q}_{\alpha_k}$. This follows immediately from general principles of convex analysis and has been recorded explicitly, for instance, in [26]. In the case of a nonatomic probability space but general mixed superquantile risk measures, we find properties of risk envelopes developed in [9, Sections 4.5 and 4.6] and [32].

For arbitrary probability spaces, the literature lacks explicit expressions for risk envelopes of mixed superquantile risk measures coming from a weighting measure λ that is not merely discrete. These risk envelopes ought, by analogy, to be a sort of "continuous sum" or integral of various sets $\mathcal{Q}_\alpha$ of the form in (9), and the contemplation of such an expression raises serious technical challenges in integration theory.

We take on those challenges here, but with some of the technical background details placed in the Appendix. Let

$$\mathcal{M} := \left\{ q : [0,1) \to \mathcal{L}^2 \ \middle| \ q \ \text{is} \ \left(\bar{\mathcal{B}}_{[0,1)}, \mathcal{B}_{\mathcal{L}^2}\right)\text{-measurable}, \ \int \|q(\beta)\|_2 \, d\lambda(\beta) < \infty \right\}.$$

Observe that $\mathcal{M}$ is well-defined because by Lemma A.5 (the "A" points to the Appendix), the mapping $\beta \mapsto \|q(\beta)\|_2$ is $\bar{\mathcal{B}}_{[0,1)}$-measurable whenever q is $(\bar{\mathcal{B}}_{[0,1)}, \mathcal{B}_{\mathcal{L}^2})$-measurable.

We are now ready to deal with the risk envelope of a mixed superquantile risk measure $\mathcal{R}$ and for this purpose utilize a collection of random variables in terms of (Bochner) integrals of elements of $\mathcal{M}$. In the following, we let $\overline{I\!R} = I\!R \cup \{-\infty, \infty\}$.

3.1 Theorem (risk envelope for mixed superquantiles). *For a mixed superquantile measure of risk $\mathcal{R}$ with associated weighting measure λ, let[7]*

$$\mathcal{Q} := \mathrm{cl} \left\{ Q \in \mathcal{L}^2 \ \middle| \ Q = \int q(\beta) \, d\lambda(\beta), q \in \mathcal{M}, q(\beta) \in \mathcal{Q}_\beta \ \text{for } \lambda\text{-a.e. } \beta \in [0,1) \right\},$$

where cl *denotes closure with respect to the (strong) topology on $\mathcal{L}^2$. Then $\mathcal{Q}$ is nonempty, convex, and is the risk envelope for $\mathcal{R}$, i.e., for any $X \in \mathcal{L}^2$,*

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} E[XQ].$$

Moreover, if $\int_0^1 1/\sqrt{1-\alpha} \, d\lambda(\alpha) < \infty$, then $\mathcal{Q}$ is also weakly compact.

Proof. Let $X \in \mathcal{L}^2$ and $f : [0,1) \times \mathcal{L}^2 \to \overline{I\!R}$ be defined by

$$f(\alpha, Q) = \begin{cases} -E[XQ] & \text{if } Q \in \mathcal{Q}_\alpha \\ \infty & \text{otherwise.} \end{cases}$$

---

[7]We note that $\mathcal{Q}$ resembles the Aumann integral (see for example [3]) of the set-valued mapping $\beta \mapsto \mathcal{Q}_\beta$.

In view of Definition A.3, f is a normal integrand because (i) f is $(\bar{\mathcal{B}}_{[0,1)} \otimes \mathcal{B}_{\mathcal{L}^2})$-measurable as the sum of the continuous[8] function $-E[X\cdot]$ on $[0,1) \times \mathcal{L}^2$ and an indicator function vanishing on the set

$$\{(\beta, Q) \in [0,1) \times \mathcal{L}^2 \mid Q \in \mathcal{Q}_\beta\} \in \bar{\mathcal{B}}_{[0,1)} \otimes \mathcal{B}_{\mathcal{L}^2}$$

and infinity elsewhere, (ii) $f(\beta, Q) \geq -E[XQ] > -\infty$ for $\beta \in [0,1)$ and $Q \in \mathcal{L}^2$, and (iii) for all $\beta \in [0,1)$, $f(\beta, \cdot)$ is lower semicontinuous by the continuity of $E[X\cdot]$ on $\mathcal{L}^2$ and the closedness of $\mathcal{Q}_\beta \subset \mathcal{L}^2$, and $f(\beta, \cdot)$ is not identical to ∞ with $Q = 1 \in \mathcal{Q}_\beta$ furnishing a finite value $f(\beta, 1) = -E[X]$. In view of Proposition A.6 and the fact that $q = 1$ provides an element of $\mathcal{M}$ with $\int f(\beta, q(\beta)) \, d\lambda(\beta) = -E[X] < \infty$, Proposition A.4 applies. Consequently, the interchange of integration and minimization is permitted and we obtain that

$$\mathcal{R}(X) = \int \sup_{Q_\beta \in \mathcal{Q}_\beta} E[XQ_\beta] \, d\lambda(\beta) = -\int \inf_{Q \in \mathcal{L}^2} f(\beta, Q) \, d\lambda(\beta)$$

$$= -\inf_{q \in \mathcal{M}} \int f(\beta, q(\beta)) \, d\lambda(\beta).$$

We next consider the interchange of integration with respect to λ and $\mathbb{P}$. For $q \in \mathcal{M}$, it follows from Lemma A.5 that the function $(\beta, \omega) \mapsto |X(\omega)q(\beta)(\omega)|$ is measurable. By Tonelli-Fubini's Theorem and Cauchy-Schwartz inequality,

$$\int |X(\omega)q(\beta)(\omega)| d(\lambda \times \mathbb{P})(\beta, \omega) = \int E[|Xq(\beta)|] \, d\lambda(\beta) \leq \|X\|_2 \int \|q(\beta)\|_2 \, d\lambda(\beta) < \infty,$$

where the finiteness follows by the property of $q \in \mathcal{M}$. Then by Tonelli-Fubini's Theorem,

$$\int E[Xq(\beta)] \, d\lambda(\beta) = E\left[X \int q(\beta) \, d\lambda(\beta)\right].$$

Since

$$\int f(\beta, q(\beta)) \, d\lambda(\beta) = \int E[Xq(\beta)] \, d\lambda(\beta)$$

whenever $q \in \mathcal{M}$ is such that $q(\beta) \in \mathcal{Q}_\beta$ for λ-a.e. $\beta \in [0,1)$ and $\int f(\beta, q(\beta)) \, d\lambda(\beta) = \infty$ otherwise, we find that

$$\inf_{q \in \mathcal{M}} \int f(\beta, q(\beta)) \, d\lambda(\beta) = \inf_{q \in \mathcal{M}} \left\{ \int E[-Xq(\beta)] \, d\lambda(\beta) + \iota(q) \right\}$$

$$= \inf_{q \in \mathcal{M}} \left\{ -E\left[X \int q(\beta) \, d\lambda(\beta)\right] + \iota(q) \right\},$$

where

$$\iota(q) = \begin{cases} 0 & \text{if } q(\beta) \in \mathcal{Q}_\beta \text{ for } \lambda\text{-a.e. } \beta \in [0,1) \\ \infty & \text{otherwise.} \end{cases}$$

Compiling the above results, we see that

$$\mathcal{R}(X) = -\inf_{q \in \mathcal{M}} \int f(\beta, q(\beta)) \, d\lambda(\beta) = \sup_{q \in \mathcal{M}} \left\{ E\left[X \int q(\beta) \, d\lambda(\beta)\right] - \iota(q) \right\} = \sup_{Q \in \mathcal{Q}} E[XQ].$$

---

[8]Here continuity is with respect to the product topology of the norm-topologies on $[0,1)$ and $\mathcal{L}^2$.

The convexity of $\mathcal{Q}$ follows from the convexity of $\mathcal{Q}_\beta$. Since $1 \in \mathcal{Q}$, $\mathcal{Q}$ is not empty. Under the additional assumption that $\int 1/\sqrt{1-\alpha}\, d\lambda(\alpha) < \infty$, $\mathcal{R}$ is finite-valued on $\mathcal{L}^2$ and even locally bounded around the origin of $\mathcal{L}^2$ by Proposition 2.3. This local boundedness for a positively homogeneous convex function, as the support function of a set $\mathcal{Q}$, corresponds to that set being bounded. Consequently, $\mathcal{Q}$ is bounded. Since $\mathcal{Q}$ is convex, weak closedness follows from strong closedness and therefore weak compactness is established. $\qquad\square$

For the special case of a second-order superquantile risk measure we then obtain the following corollary.

3.2 Corollary (risk envelope for second-order superquantiles). *For $\alpha \in [0,1)$, the risk envelope for $\bar{\mathcal{R}}_\alpha$ is given by*

$$\bar{\mathcal{Q}}_\alpha := \mathrm{cl} \left\{ Q \in \mathcal{L}^2 \;\middle|\; Q = \frac{1}{1-\alpha} \int_\alpha^1 q(\beta)d\beta, q \in \mathcal{M},\ q(\beta) \in \mathcal{Q}_\beta \text{ for } m\text{-a.e. } \beta \in [\alpha,1) \right\}.$$

Moreover, $\bar{\mathcal{Q}}_\alpha$ is a nonempty weakly-compact convex subset of $\mathcal{L}^2$.

In addition to the trivial cases when λ and/or $\mathbb{P}$ are positive only on a finite number of points in $[0,1)$ and Ω, respectively, the closure in the definition of $\mathcal{Q}$ is unnecessary under the following condition.

3.3 Proposition (dispensing with the closure operation). *Suppose that λ is nonatomic and also that $\int_0^1 1/(1-\alpha)\, d\lambda(\alpha) < \infty$. Then the closure operation is superfluous in the expression of the envelope in Theorem 3.1. One can simply take*

$$\mathcal{Q} = \left\{ Q \in \mathcal{L}^2 \;\middle|\; Q = \int q(\beta)\, d\lambda(\beta), q \in \mathcal{M}, q(\beta) \in \mathcal{Q}_\beta \text{ for } \lambda\text{-a.e. } \beta \in [0,1) \right\}.$$

Proof. By [6], an integrably bounded $\bar{\mathcal{B}}_{[0,1)}$-measurable set-valued mapping $S : [0,1) \rightrightarrows \mathcal{L}^2$, with closed and convex values, satisfies

$$\mathrm{cl} \left\{ \int S(\alpha)\, d\lambda(\alpha) \right\} = \int S(\alpha)\, d\lambda(\alpha)$$

when λ is nonatomic. Take S to be the mapping $\alpha \mapsto \{q(\alpha) \mid q \in \mathcal{M}, q(\alpha) \in \mathcal{Q}_\alpha\}$, which obviously is closed and convex valued by the properties of $\mathcal{Q}_\alpha$. Moreover, since both $[0,1)$ and $\mathcal{L}^2$ are separable, there exists a countable collection $\{q^i\}_{i=1}^\infty$, $q^i \in \mathcal{M}$, such that $S(\alpha) = \mathrm{cl}\{q^i(\alpha) \mid i = 1,2,...\}$ for λ-a.e. $\alpha \in [0,1)$. Thus, S is $\bar{\mathcal{B}}_{[0,1)}$-measurable; see for example [18, Theorem 1]. The mapping S is integrably bounded if there exists a $\bar{\mathcal{B}}_{[0,1)}$-measurable $g : [0,1) \to \mathbb{R}$ with $\int g(\alpha)\, d\lambda(\alpha) < \infty$ and

$$\sup_{Q \in S(\alpha)} \|Q\|_2 \le g(\alpha) \quad \text{for } \lambda\text{-a.e. } \alpha \in [0,1).$$

Since for our choice of S we have that every $Q \in S(\alpha)$ has $Q(\omega) \le 1/(1-\alpha)$ for a.e. $\omega \in \Omega$, integrably boundedness holds with $g(\alpha) = 1/(1-\alpha)$ under the imposed restriction on λ. $\qquad\square$

Next, we turn to specific expressions for *risk identifiers*. Recall from (7) that for any $X \in \mathcal{L}^2$ and positively homogeneous regular measure of risk on $\mathcal{L}^2$, a Q in the risk envelope of the risk measure that maximizes $E[XQ]$ is called a risk identifier at X. We again start with the building blocks from first-order superquantile risk measures.

For $X \in \mathcal{L}^2$, the set

$$\mathcal{Q}_\alpha^X := \underset{Q \in \mathcal{Q}_\alpha}{\operatorname{argmax}} E[XQ]$$

is convex and nonempty with its elements referred to as risk identifiers of $\mathcal{R}_\alpha$. Before we characterize these risk identifiers, we introduce additional notation.

For $\beta \in (0,1)$, let

$$\Omega_\beta(X) := \{\omega \in \Omega \mid X(\omega) = q_X(\beta)\}$$

and let

$$F_X^-(x) := \lim_{x' \nearrow x} F_X(x'), \quad x \in \mathbb{R}$$

be the left-continuous "companion" of the cumulative distribution function F_X, where the limit exists by the virtue of F_X being nondecreasing and bounded from above. For F_X continuous, $F_X = F_X^-$ of course.

The risk identifiers of $\mathcal{R}_\alpha$ are then characterized as follows; see also [31, Equation 4.21] for closely related expressions.

3.4 Proposition *For $X \in \mathcal{L}^2$ and $\beta \in (0,1)$, let $r_\beta^X \in \mathcal{L}^2$ be such that*

$$0 \leq r_\beta^X(\omega) \leq \frac{1}{1-\beta} \text{ for a.e. } \omega \in \Omega \text{ and } \int_{\Omega_\beta(X)} r_\beta^X(\omega) d\mathbb{P}(\omega) = \frac{F_X(q_X(\beta)) - \beta}{1-\beta}. \tag{12}$$

Every such r_β^X, defines a unique[9] $Q_\beta^{X,r_\beta^X} \in \mathcal{L}^2$ given for a.e. $\omega \in \Omega$ by

$$Q_\beta^{X,r_\beta^X}(\omega) := \begin{cases} \frac{1}{1-\beta} & \text{if } X(\omega) > q_X(\beta) \\ r_\beta^X(\omega) & \text{if } X(\omega) = q_X(\beta) \text{ and } \mathbb{P}(\{\omega\}) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Then,

$$\mathcal{Q}_\beta^X = \left\{ Q \in \mathcal{L}^2 \,\middle|\, Q = Q_\beta^{X,r_\beta^X} \text{ for some } r_\beta^X \in \mathcal{L}^2 \text{ satisfying (12)} \right\}.$$

Moreover,

$$\mathcal{Q}_0^X = \{Q \in \mathcal{L}^2 \mid Q(\omega) = 1 \text{ for a.e. } \omega \in \Omega\}.$$

Proof. Let $\beta \in (0,1)$ and $X \in \mathcal{L}^2$. We first show that there exists an $r_\beta^X \in \mathcal{L}^2$ satisfying (12). For $\omega \in \Omega$ satisfying $X(\omega) = q_X(\beta)$ and $\mathbb{P}(\{\omega\}) > 0$, $F_X^-(X(\omega)) \leq \beta \leq F_X(X(\omega))$, with at least one of the inequalities being strict, and

$$\frac{F_X(X(\omega)) - \beta}{(1-\beta)(F_X(X(\omega)) - F_X^-(X(\omega)))} \in [0, 1/(1-\beta)].$$

---

[9]With $\mathcal{L}^2$ consisting of equivalence classes of functions identical up to on a set of $\mathbb{P}$-measure zero, uniqueness of course is in the sense of such equivalence classes.

11

Let $\hat{r}_\beta^X \in \mathcal{L}^2$ be defined for a.e. $\omega \in \Omega$ by

$$\hat{r}_\beta^X(\omega) := \begin{cases} \frac{F_X(X(\omega))-\beta}{(1-\beta)(F_X(X(\omega))-F_X^-(X(\omega)))}, & \text{if } X(\omega) = q_X(\beta) \text{ and } \mathbb{P}(\{\omega\}) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Clearly, $\hat{r}_\beta^X$ satisfies $0 \leq \hat{r}_\beta^X(\omega) \leq 1/(1-\beta)$ for a.e. $\omega \in \Omega$. Moreover,

$$\int_{\Omega_\beta(X)} \hat{r}_\beta^X(\omega)d\mathbb{P}(\omega) = \int_{\Omega_\beta(X)} \frac{F_X(q_X(\beta))-\beta}{(1-\beta)(F_X(q_X(\beta))-F_X^-(q_X(\beta)))}d\mathbb{P}(\omega) = \frac{F_X(q_X(\beta))-\beta}{1-\beta}$$

and $\hat{r}_\beta^X$ therefore satisfies (12).

Let $r_\beta^X \in \mathcal{L}^2$ satisfy (12). Since $0 \leq Q_\beta^{X,r_\beta^X}(\omega) \leq 1/(1-\beta)$ for a.e. $\omega \in \Omega$ and

$$\int Q_\beta^{X,r_\beta^X}(\omega)d\mathbb{P}(\omega) = \int_{\{\omega \in \Omega \mid X(\omega) > q_X(\beta)\}} \frac{1}{1-\beta}d\mathbb{P}(\omega) + \int_{\Omega_\beta(X)} r_\beta^X(\omega)d\mathbb{P}(\omega)$$

$$= \frac{1-F_X(q_X(\beta))}{1-\beta} + \frac{F_X(q_X(\beta))-\beta}{1-\beta} = 1,$$

we find that $Q_\beta^{X,r_\beta^X} \in \mathcal{Q}_\beta$. Moreover,

$$E\left[XQ_\beta^{X,r_\beta^X}\right] = \int_{\{\omega \in \Omega \mid X(\omega) > q_X(\beta)\}} \frac{X(\omega)}{1-\beta}d\mathbb{P}(\omega) + \int_{\Omega_\beta(X)} X(\omega)r_\beta^X(\omega)d\mathbb{P}(\omega)$$

$$= \frac{1}{1-\beta} \int_{\{\omega \in \Omega \mid X(\omega) > q_X(\beta)\}} X(\omega)d\mathbb{P}(\omega) + q_X(\beta)\frac{F_X(q_X(\beta))-\beta}{1-\beta}$$

$$= \int_{-\infty}^\infty x \, dF_X^\beta(x),$$

where

$$F_X^\beta(x) := \begin{cases} \frac{F_X(x)-\beta}{1-\beta} & \text{if } F_X(x) \geq \beta \\ 0 & \text{if } F_X(x) < \beta. \end{cases}$$

It is well known (see [25]) that the superquantile $\bar{q}_X(\beta) = \int_{-\infty}^\infty x \, dF_X^\beta(x)$. Thus, we have proved that Q_β^{X,r_β^X} maximizes $E[X\cdot]$ over $\mathcal{Q}_\beta$. Any $Q \in \mathcal{Q}_\beta$ not equal to Q_β^{X,r_β^X} for any r_β^X must necessarily have $E[XQ] < \bar{q}_X(\beta)$.

The case of $\beta = 0$ follows also as then $\mathcal{Q}_0 = \{Q \in \mathcal{L}^2 \mid 0 \leq Q(\omega) \leq 1 \text{ for a.e. } \omega \in \Omega, E[Q] = 1\}$. $\square$

A particular element of $\mathcal{Q}_\beta^X$ plays a central role in the following. Let $\hat{r}_\beta^X \in \mathcal{L}^2$ be as defined in (14). Consequently by Proposition 3.4, $\hat{Q}_\beta^X$ defined for a.e. $\omega \in \Omega$ by

$$\hat{Q}_\beta^X(\omega) := \begin{cases} \frac{1}{1-\beta} & \text{if } X(\omega) > q_X(\beta) \\ \hat{r}_\beta^X(\omega) & \text{if } X(\omega) = q_X(\beta) \text{ and } \mathbb{P}(\{\omega\}) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

12

is a point in $\mathcal{Q}_\beta^X$. Moreover, let $\hat{Q}_0^X \in \mathcal{L}^2$ be defined by $\hat{Q}_0^X(\omega) = 1$ for a.e. $\omega \in \Omega$, which therefore by Proposition 3.4 is a point in $\mathcal{Q}_0^X$. The random variable $\hat{Q}_\beta^X$ behaves continuously in β in a sense given next.

3.5 Proposition *If* $\beta^\nu, \beta \in [0, 1)$ *and* $\beta^\nu \to \beta$, *then for any* $X \in \mathcal{L}^2$, $\|\hat{Q}_{\beta^\nu}^X - \hat{Q}_\beta^X\|_2 \to 0$.

Proof. Let $X \in \mathcal{L}^2$ and $\hat{r}_\beta^X$ be defined in (14) and $\beta \in (0, 1)$. Suppose that $F_X(q_X(\beta)) - F_X^-(q_X(\beta)) > 0$. We consider two cases.

First, suppose that $\beta^\nu \to \beta$, with $\beta^\nu < \beta$ for all ν, which implies that $\beta \in [F_X^-(q_X(\beta)), F_X(q_X(\beta))]$. If $\beta \in (F_X^-(q_X(\beta)), F_X(q_X(\beta))]$, then $q_X(\beta^\nu) = q_X(\beta)$ for sufficiently large ν. Consequently, for sufficiently large ν,

$$\|\hat{Q}_{\beta^\nu}^X - \hat{Q}_\beta^X\|_2^2 = \int_{\{\omega | X(\omega) < q_X(\beta)\}} (0 - 0)^2 d\mathbb{P}(\omega)$$
$$+ \int_{\Omega_\beta(X)} (\hat{r}_{\beta^\nu}^X(\omega) - \hat{r}_\beta^X(\omega))^2 d\mathbb{P}(\omega) + \int_{\{\omega | X(\omega) > q_X(\beta)\}} \left(\frac{1}{1 - \beta^\nu} - \frac{1}{1 - \beta} \right)^2 d\mathbb{P}(\omega).$$

When $X(\omega) = q_X(\beta^\nu) = q_X(\beta)$,

$$\hat{r}_{\beta^\nu}^X(\omega) - \hat{r}_\beta^X(\omega) = \frac{F_X(q_X(\beta)) - \beta^\nu}{(1 - \beta^\nu)(F_X(q_X(\beta)) - F_X^-(q_X(\beta)))} - \frac{F_X(q_X(\beta)) - \beta}{(1 - \beta)(F_X(q_X(\beta)) - F_X^-(q_X(\beta)))}$$

Hence, all three terms in the above integral vanish as $\nu \to \infty$. If $\beta = F_X^-(q_X(\beta))$, then we only have that $q_X(\beta^\nu) \nearrow q_X(\beta)$ by the left-continuity of q_X and in fact $q_X(\beta^\nu) < q_X(\beta))$ for all ν. Consequently,

$$\|\hat{Q}_{\beta^\nu}^X - \hat{Q}_\beta^X\|_2^2 = \int_{\{\omega | X(\omega) < q_X(\beta^\nu)\}} (0 - 0)^2 d\mathbb{P}(\omega)$$
$$+ \int_{\{\omega | q_X(\beta^\nu) = X(\omega) < q_X(\beta)\}} (\hat{r}_{\beta^\nu}^X(\omega) - 0)^2 d\mathbb{P}(\omega)$$
$$+ \int_{\{\omega | q_X(\beta^\nu) < X(\omega) = q_X(\beta)\}} \left(\frac{1}{1 - \beta^\nu} - \hat{r}_\beta^X(\omega) \right)^2 d\mathbb{P}(\omega)$$
$$+ \int_{\{\omega | q_X(\beta^\nu) < q_X(\beta) < X(\omega)\}} \left(\frac{1}{1 - \beta^\nu} - \frac{1}{1 - \beta} \right)^2 d\mathbb{P}(\omega).$$

Of the four integrals, the first and fourth ones obviously tend to zero. For the second one, we see that

$$\mathbb{P}(\{\omega | q_X(\beta^\nu) < X(\omega) = q_X(\beta)\}) = F_X(q_X(\beta^\nu)) - F_X^-(q_X(\beta^\nu)) \leq F_X(q_X(\beta)) - F_X^-(q_X(\beta^\nu)) \to 0$$

by the left-continuity of F_X^- and consequently the integral also tends to zero. For the third integral, we find that when $X(\omega) = q_X(\beta)$

$$\hat{r}_\beta^X(\omega) = \frac{F_X(q_X(\beta)) - \beta}{(1 - \beta)(F_X(q_X(\beta)) - F_X^-(q_X(\beta)))} = \frac{F_X(q_X(\beta)) - F_X^-(q_X(\beta))}{(1 - \beta)(F_X(q_X(\beta)) - F_X^-(q_X(\beta)))} = \frac{1}{1 - \beta}.$$

Consequently, the third integral also tends to zero.

Second, suppose that $\beta^\nu \to \beta$, with $\beta^\nu > \beta$ for all ν. If $\beta \in [F_X^-(q_X(\beta)), F_X(q_X(\beta)))$, then $q_X(\beta^\nu) = q_X(\beta)$ for sufficiently large ν and the corresponding argument for the first case still holds. If $\beta = F_X(q_X(\beta))$, then we only have that $q_X(\beta^\nu) > q_X(\beta))$ for all ν. Consequently,

$$
\begin{aligned}
\|\hat{Q}_{\beta^\nu}^X - \hat{Q}_\beta^X\|_2^2 = & \int_{\{\omega | X(\omega) < q_X(\beta)\}} (0-0)^2 d\mathbb{P}(\omega) \\
& + \int_{\{\omega | q_X(\beta) = X(\omega) < q_X(\beta^\nu)\}} (0 - \hat{r}_\beta^X(\omega))^2 d\mathbb{P}(\omega) \\
& + \int_{\{\omega | q_X(\beta) < X(\omega) = q_X(\beta^\nu)\}} \left(\hat{r}_{\beta^\nu}^X(\omega) - \frac{1}{1-\beta} \right)^2 d\mathbb{P}(\omega) \\
& + \int_{\{\omega | q_X(\beta) < q_X(\beta^\nu) < X(\omega)\}} \left(\frac{1}{1-\beta^\nu} - \frac{1}{1-\beta} \right)^2 d\mathbb{P}(\omega).
\end{aligned}
$$

The first and fourth integrals obviously tend to zero. For the second one,

$$
\hat{r}_\beta^X(\omega) = \frac{F_X(q_X(\beta)) - \beta}{(1-\beta)(F_X(q_X(\beta)) - F_X^-(q_X(\beta)))} = \frac{F_X(q_X(\beta)) - F_X(q_X(\beta))}{(1-\beta)(F_X(q_X(\beta)) - F_X^-(q_X(\beta)))} = 0
$$

and consequently a zero integral. For the third integral,

$$
\hat{r}_{\beta^\nu}^X(\omega) = \frac{F_X(q_X(\beta^\nu)) - \beta^\nu}{(1-\beta^\nu)(F_X(q_X(\beta^\nu)) - F_X^-(q_X(\beta^\nu)))} \to \frac{1}{1-\beta}
$$

if $q_X(\beta^\nu)$ remains bounded away from $q_X(\beta)$ because then $F_X^-(\beta^\nu) \to F_X(\beta) = \beta$. If $q_X(\beta^\nu) \to q_X(\beta)$, then by the right-continuity of F_X we have that

$$
\mathbb{P}(\{\omega \in \Omega | q_X(\beta) < X(\omega) = q_X(\beta^\nu)\}) = F_X(q_X(\beta^\nu)) - F_X^-(q_X(\beta^\nu)) \leq F_X(q_X(\beta^\nu)) - F_X(q_X(\beta)) \to 0.
$$

Consequently, the third integral also tends to zero.

The situation with $F_X(q_X(\beta)) - F_X^-(q_X(\beta)) = 0$ follows with similar and in fact simplified arguments as in that case F_X is continuous at $q_X(\beta)$ and q_X is continuous at β.

Finally, we consider the case with $\beta = 0$ and $\beta^\nu \searrow 0$. Then,

$$
\begin{aligned}
\|\hat{Q}_{\beta^\nu} - \hat{Q}_0\|_2^2 = & \int_{\{\omega | X(\omega) > q_X(\beta^\nu)\}} \left(\frac{1}{1-\beta^\nu} - 1 \right)^2 d\mathbb{P}(\omega) \\
& + \int_{\Omega_{\beta^\nu}(X)} (\hat{r}_{\beta^\nu}(\omega) - 1)^2 d\mathbb{P}(\omega) + \int_{\{\omega | X(\omega) < q_X(\beta^\nu)\}} (0 - 1)^2 d\mathbb{P}(\omega).
\end{aligned}
$$

Since $1/(1-\beta^\nu) \to 1$, the first integral vanishes. The last two integrals vanish since their integrands are bounded and $F_X(q_X(\beta^\nu)) \to 0$. $\square$

We are then in a position to characterize risk identifiers of mixed superquantile risk measures. For $X \in \mathcal{L}^2$, let

$$
\mathcal{Q}^X := \mathrm{cl}\left\{ Q \in \mathcal{L}^2 \;\middle|\; Q = \int q(\beta)\, d\lambda(\beta), q \in \mathcal{M}, q(\beta) \in \mathcal{Q}_\beta^X \text{ for } \lambda\text{-a.e. } \beta \in [0,1) \right\}. \tag{16}
$$

14

3.6 Theorem (risk identifiers for mixed superquantiles). *For $X \in \mathcal{L}^2$, the set $\mathcal{Q}^X$ is convex and satisfies the following.*

(i) *If $Q \in \mathcal{Q}^X$, then Q is a risk identifier of $\mathcal{R}$ at X.*

(ii) *If $\int 1/\sqrt{1-\beta}\, d\lambda(\beta) < \infty$, then $\mathcal{Q}^X$ is nonempty and weakly compact, and $Q \in \mathcal{Q}^X$ whenever Q is a risk identifier of $\mathcal{R}$ at X. Moreover, $\hat{Q} := \int \hat{q}(\beta)\, d\lambda(\beta)$, where*

$$\hat{q} : [0,1) \to \mathcal{L}^2, \text{ with } \hat{q}(\beta) = \hat{Q}^X_\beta \text{ (defined in (15)) for all } \beta \in [0,1),$$

is furnishing an element of $\mathcal{Q}^X$.

Proof. We first consider (i). Let $Q \in \mathcal{Q}^X$. There exists sequences $\{Q^\nu\}_{\nu=1}^\infty \subset \mathcal{L}^2$ and $\{q^\nu\}_{\nu=1}^\infty \subset \mathcal{M}$ such that $\|Q^\nu - Q\|_2 \to 0$, $Q^\nu = \int q^\nu(\beta)\, d\lambda(\beta)$, and $q^\nu(\beta) \in \mathcal{Q}^X_\beta$ for all ν and λ-a.e. $\beta \in [0,1)$. Then, for every ν,

$$\mathcal{R}(X) = \int E[Xq^\nu(\beta)]\, d\lambda(\beta) = E\left[X \int q^\nu(\beta)\, d\lambda(\beta)\right] = E[XQ^\nu],$$

where the middle equality follows by the same argument as in the proof of Theorem 3.1. Since by the Cauchy-Schwartz inequality $E[XQ^\nu] \to E[XQ]$, we also have that $\mathcal{R}(X) = E[XQ]$, which establishes (i).

Next, we consider (ii). Suppose that $\int 1/\sqrt{1-\beta}\, d\lambda(\beta) < \infty$. We proceed toward a contradiction. Suppose that $Q \in \mathcal{Q}$ is a risk identifier of $\mathcal{R}$ at X, but $Q \notin \mathcal{Q}^X$. Then there must exists a $q \in \mathcal{M}$ and $B \in \bar{\mathcal{B}}_{[0,1)}$ such that $q(\beta) \in \mathcal{Q}_\beta$ for λ-a.e. $\beta \in [0,1)$, $\lambda(B) > 0$, and $q(\beta) \notin \mathcal{Q}^X_\beta$ for all $\beta \in B$. However, this implies that $E[Xq(\beta)] < E[XQ^X_\beta]$ for all $\beta \in B$ and any $Q^X_\beta \in \mathcal{Q}^X_\beta$. Consequently, $E[XQ] < \mathcal{R}(X)$, which is a contradiction.

Since $\mathcal{Q}$ is weakly compact by Theorem 3.1, the weak compactness of $\mathcal{Q}^X$ follows from it being a closed convex subset of $\mathcal{Q}$. Finally, we show that $\hat{Q} \in \mathcal{Q}^X$. The conclusion follows when we have shown that $\hat{q} \in \mathcal{M}$. By Proposition 3.5, $\hat{q}$ is continuous and therefore $(\bar{\mathcal{B}}_{[0,1)}, \mathcal{B}_{\mathcal{L}^2})$-measurable. Since for $\beta \in (0,1)$

$$\|\hat{Q}^X_\beta\|^2_2 = \int_{\{\omega \in \Omega \; | \; X(\omega) > q_X(\beta)\}} \frac{1}{(1-\beta)^2} d\mathbb{P}(\omega) + \int_{\Omega_\beta(X)} \left[\frac{F_X(q_X(\beta)) - \beta}{(1-\beta)(F_X(q_X(\beta)) - F^-_X(q_X(\beta)))}\right]^2 d\mathbb{P}(\omega)$$

$$= \frac{1-\beta}{(1-\beta)^2} + \left[\frac{F_X(q_X(\beta)) - \beta}{(1-\beta)(F_X(q_X(\beta)) - F^-_X(q_X(\beta)))}\right]^2 (F_X(q_X(\beta)) - F^-_X(q_X(\beta)))$$

$$= \frac{1}{1-\beta} + \frac{(F_X(q_X(\beta)) - \beta)^2}{(1-\beta)^2(F_X(q_X(\beta)) - F^-_X(q_X(\beta)))}$$

$$\leq \frac{1}{1-\beta} + \frac{(1-\beta)(F_X(q_X(\beta)) - \beta)}{(1-\beta)^2(F_X(q_X(\beta)) - F^-_X(q_X(\beta)))}$$

$$\leq \frac{1}{1-\beta} + \frac{F_X(q_X(\beta)) - F^-_X(q_X(\beta))}{(1-\beta)(F_X(q_X(\beta)) - F^-_X(q_X(\beta)))} = \frac{2}{1-\beta}$$

and $\|\hat{Q}^X_0\|^2_2 = 1$, we find that

$$\int \|\hat{q}(\beta)\|_2\, d\lambda(\beta) \leq \sqrt{2} \int \frac{1}{\sqrt{1-\beta}}\, d\lambda(\beta) < \infty.$$

15

Consequently $\hat{q} \in \mathcal{M}$ and $\hat{Q} = \int \hat{q}(\beta)\,d\lambda(\beta) \in \mathcal{Q}^X$, which complete the proof. $\qquad\square$

We observe that when $\int 1/\sqrt{1-\beta}\,d\lambda(\beta) = \infty$, there are random variables $X \in \mathcal{L}^2$ with $\mathcal{R}(X) = \infty$. In this case it might not be necessary to select q in (16) with $q(\beta) \in \mathcal{Q}_\beta^X$ for λ-a.e. $\beta \in [0,1)$ because $\int E[Xq(\beta)]\,d\lambda(\beta)$ might still be infinity. For the special case of a second-order superquantile risk measure, we directly obtain the following corollary without this complication.

3.7 Corollary *For $\alpha \in [0,1)$ and $X \in \mathcal{L}^2$, the set*

$$\bar{\mathcal{Q}}_\alpha^X := \mathrm{cl}\left\{ Q \in \mathcal{L}^2 \;\middle|\; Q = \frac{1}{1-\alpha}\int_\alpha^1 q(\beta)d\beta, q \in \mathcal{M}, q(\beta) \in \mathcal{Q}_\beta^X \text{ for } m\text{-a.e. } \beta \in [\alpha,1) \right\}$$

is nonempty, convex, and weakly compact. Moreover,

$$Q \in \bar{\mathcal{Q}}_\alpha^X \text{ if and only if } Q \text{ is a risk identifier of } \mathcal{R}_\alpha \text{ at } X.$$

$\qquad\square$

Further simplifications are possible in the case of second-order superquantile risk measures. As usual, we interpret 0 times $-\infty$ as zero in the following.

3.8 Theorem (further characterization of second-order superquantile risk identifiers). *For $X \in \mathcal{L}^2$ and $\alpha \in [0,1)$, $\bar{\mathcal{Q}}_\alpha^X$ is the closure of elements $\bar{Q}_\alpha^X \in \bar{\mathcal{Q}}_\alpha$ given, for a.e. $\omega \in \Omega$, by*

$$\bar{Q}_\alpha^X(\omega) = \begin{cases} \frac{1}{1-\alpha}\left[\log\frac{1-\alpha}{1-f(\omega)} + \int_{f(\omega)}^{F(\omega)} r_\beta^X(\omega)d\beta\right] & \text{if } \alpha < f(\omega) < 1 \\ \frac{1}{1-\alpha}\int_\alpha^{F(\omega)} r_\beta^X(\omega)d\beta & \text{if } f(\omega) \le \alpha \le F(\omega) \\ 0 & \text{otherwise,} \end{cases}$$

where $r_\beta^X \in \mathcal{L}^2$ satisfies (12) and $F(\omega) := F_X(X(\omega))$ and $f(\omega) := F_X^-(X(\omega))$.

The specific choice $\hat{r}_\beta^X \in \mathcal{L}^2$ given in (14) results in the risk identifier $\bar{Q}_\alpha^X \in \bar{\mathcal{Q}}_\alpha^X$ having, for a.e. $\omega \in \Omega$,

$$\bar{Q}_\alpha^X(\omega) = \begin{cases} \frac{1}{1-\alpha}\log\frac{1-\alpha}{1-F(\omega)} & \text{if } \alpha < f(\omega) = F(\omega) < 1 \\ \frac{1}{1-\alpha}\left[\log\frac{1-\alpha}{1-f(\omega)} + 1 + \frac{1-F(\omega)}{F(\omega)-f(\omega)}\log\frac{1-F(\omega)}{1-f(\omega)}\right] & \text{if } \alpha < f(\omega) < F(\omega) \\ \frac{1}{1-\alpha}\left[\frac{F(\omega)-\alpha}{F(\omega)-f(\omega)} + \frac{1-F(\omega)}{F(\omega)-f(\omega)}\log\frac{1-F(\omega)}{1-\alpha}\right] & \text{if } f(\omega) \le \alpha \le F(\omega) \text{ and } f(\omega) < F(\omega) \\ 0 & \text{otherwise.} \end{cases}$$

Proof. For $\omega \in \Omega$ such that $\alpha < F_X^-(X(\omega)) < 1$,

$$\int_{\{\beta \in (\alpha,1) \;|\; X(\omega) > q_X(\beta)\}} \frac{1}{1-\beta}d\beta = [-\log(1-\beta)]_\alpha^{F_X^-(X(\omega))} = \log\frac{1-\alpha}{1-F_X^-(X(\omega))}. \tag{17}$$

By Proposition 3.4,

$$\bar{Q}_\alpha^X(\omega) = \frac{1}{1-\alpha}\left[\int_{\{\beta \in (\alpha,1) \;|\; X(\omega) > q_X(\beta)\}} \frac{1}{1-\beta}d\beta + \int_{\{\beta \in (\alpha,1) \;|\; X(\omega) = q_X(\beta)\}} r_\beta^X(\omega)d\beta\right]$$

$$= \frac{1}{1-\alpha}\left[\log\frac{1-\alpha}{1-F_X^-(X(\omega))} + \int_{F_X^-(X(\omega))}^{F_X(X(\omega))} r_\beta^X(\omega)d\beta\right],$$

16

which proves the first claim. The second claim follows by a similar argument.

We next turn to the specific choice of $\hat{r}_\beta^X$. For $\alpha < F_X^-(X(\omega)) = F_X(X(\omega)) < 1$, the conclusion follows trivially. For $\alpha < F_X^-(X(\omega)) < F_X(X(\omega))$, integration gives that

$$\int_{F_X^-(X(\omega))}^{F_X(X(\omega))} \hat{r}_\beta^X(\omega)d\beta = \int_{F_X^-(X(\omega))}^{F_X(X(\omega))} \frac{F_X(X(\omega)) - \beta}{(1-\beta)(F_X(X(\omega)) - F_X^-(X(\omega)))}d\beta$$

$$= 1 + \frac{1 - F_X(X(\omega))}{F_X(X(\omega)) - F_X^-(X(\omega))} \log \frac{1 - F_X(X(\omega))}{1 - F_X^-(X(\omega))},$$

and the corresponding conclusion follows. The last case follows by a similar calculation. $\qquad\square$

The situation is especially simple for the following case.

3.9 Corollary *Suppose that F_X is continuous for $X \in \mathcal{L}^2$ and $\alpha \in [0,1)$. Then, $\bar{\mathcal{Q}}_\alpha^X$ is a singleton[10] with element given, for a.e. $\omega \in \Omega$, by*

$$\bar{Q}_\alpha^X(\omega) = \begin{cases} \frac{1}{1-\alpha} \log \frac{1-\alpha}{1-F_X(X(\omega))} & \text{if } \alpha < F_X(X(\omega)) < 1 \\ 0 & \text{otherwise.} \end{cases}$$

$\qquad\square$

It is obvious that expressions of risk identifiers provide alternative expressions for risk measures. Specifically, for $X \in \mathcal{L}^2$,

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} \int X(\omega)Q(\omega)dP(\omega) = \int X(\omega)Q^X(\omega)dP(\omega),$$

for any $Q^X \in \mathcal{Q}^X$. In the case of the previous corollary, it is easy to see that the second-order superquantile risk takes the simple form

$$\bar{\mathcal{R}}_\alpha(X) = \frac{1}{1-\alpha} \int_{q_X(\alpha)}^\infty x \log \frac{1-\alpha}{1-F_X(x)}dF_X(x),$$

where $q_X(\alpha) = -\infty$ for $\alpha = 0$, which complements the expression of Theorem 2.4.

4 Applications to Optimization and Regression

In applications arising in optimization under uncertainty and generalized regression, one is not only interested in the risk of a single random variable X, but rather of a parameterized family of random variables over which the "best" is to be selected according to some criterion and constraints. When the criterion and/or the constraints are given in terms of measures of risk applied to this family of random variables, we obtain optimization problems involving *parameterized risk*. Properties of these measures of risk as functions of the parameters as well as formulae for the functions' (sub)gradients become central. In this section, we discuss optimization problems involving parameterized mixed and

---

[10]Again, uniqueness is up to on a set of $\mathbb{P}$-measure zero.

second-order superquantile risk. In particular, we see that risk identifiers developed in this paper are central for expressions of subgradients.

We consider a family of random variables $X_u = g(u, \cdot)$, $u \in I\!\!R^n$, generated by the function $g : I\!\!R^n \times \Omega \to \overline{I\!\!R}$. Consistent with the previous sections, we assume that $X_u \in \mathcal{L}^2$ for all $u \in I\!\!R^n$. For a weighting measure λ and the corresponding mixed superquantile risk measure $\mathcal{R}$, as before given by

$$\mathcal{R}(X_u) = \int \bar{q}_{X_u}(\beta) \; d\lambda(\beta),$$

we get a function

$$f(u) := \mathcal{R}(X_u), \quad u \in I\!\!R^n, \tag{18}$$

representing parameterized risk. One might then proceed with determining a $u \in I\!\!R^n$ that

$$\text{minimizes } f(u) \text{ over a subset of } I\!\!R^n$$

or, alternatively, with determining a $u \in I\!\!R^n$ that

$$\text{minimizes some criterion function of } u \text{ subject to } f(u) \leq 0 \text{ and possibly other constraints.}$$

Algorithms such as cutting plane and bundle methods for solving these optimization problems require expressions for (sub)gradients of f. Justification for these approaches is provided by the Convexity Theorem of [26], which establishes that f is convex whenever $g(\cdot, \omega)$ is convex for a.e. $\omega \in \Omega$.

In the remainder of the paper, we give expressions for subgradients of f, but refrain from discussing full algorithms; see for example [16, 13, 30] for risk minimization algorithms based on dual approaches and [31] for related subgradient expressions. However, we end the paper with a discussion of primal and dual methods in the context of superquantile regression.

4.1 Subgradients of Parameterized Risk

We restrict the attention to the case with $\int 1/\sqrt{1 - \alpha} \; d\lambda(\alpha) < \infty$ which ensures the finiteness of $\mathcal{R}$ on $\mathcal{L}^2$ and also the weak compactness of $\mathcal{Q}$. We equip $I\!\!R^n \times \mathcal{L}^2$ with the product topology generated by the norm topology on $I\!\!R^n$ and the weak topology on $\mathcal{L}^2$. The convergence of points in $I\!\!R^n \times \mathcal{L}^2$ in this weak sense is denoted by $\to^w$.

For notational convenience, we let $h : I\!\!R^n \times \mathcal{L}^2 \to \overline{I\!\!R}$ be given by

$$h(u, Q) := \int g(u, \omega) Q(\omega) d\mathbb{P}(\omega). \tag{19}$$

Properties of this function are recorded next. The development here is rather standard (see for example [33]), but is included as it explicitly shows the central role of risk identifiers.

4.1 Proposition *Consider h in (19) and suppose for an open set $U \subset I\!\!R^n$ that*

(i) there exists an $L \in \mathcal{L}^2$ such that

$$|g(u, \omega) - g(u', \omega)| \leq L(\omega) \|u - u'\| \text{ for all } u, u' \in U \text{ and a.e. } \omega \in \Omega$$

(ii) for every $i = 1,...,n$, there exists an $\Omega_i \subset \Omega$, with $\mathbb{P}\{\Omega_i\} = 1$, and an $L_i \in \mathcal{L}^2$ such that $\partial g(u,\omega)/\partial u_i$ exists for $u \in U$ and $\omega \in \Omega_i$, and

$$\left| \frac{\partial g(u,\omega)}{\partial u_i} - \frac{\partial g(u',\omega)}{\partial u_i} \right| \leq L_i(\omega)\|u - u'\| \text{ for all } u, u' \in U \text{ and } \omega \in \Omega_i$$

(iii) $g(v,\cdot), \partial g(v^i,\cdot)/\partial u_i \in \mathcal{L}^2$ for some $v, v^i \in U$, $i = 1,...,n$.

Then, h is weakly continuous on $U \times \mathcal{L}^2$ and $\nabla_u h$ exists and is likewise weakly continuous on $U \times \mathcal{L}^2$.

Proof. First we consider h, which is well-defined and finite on $U \times \mathcal{L}^2$ from assumptions (i) and (iii). Suppose that $(u^\nu, Q^\nu) \to^w (u, Q)$, with $u^\nu, u \in U$ and $Q^\nu, Q \in \mathcal{L}^2$. Then by the triangle and Cauchy-Schwartz inequalities and assumption (i),

$$|h(u^\nu, Q^\nu) - h(u, Q)| \leq \left| \int [g(u^\nu,\omega) - g(u,\omega)]Q^\nu(\omega)d\mathbb{P}(\omega) \right| + \left| \int g(u,\omega)[Q^\nu(\omega) - Q(\omega)]d\mathbb{P}(\omega) \right|$$

$$\leq \|g(u^\nu,\cdot) - g(u,\cdot)\|_2 \|Q^\nu\|_2 + \left| \int g(u,\omega)[Q^\nu(\omega) - Q(\omega)]d\mathbb{P}(\omega) \right|$$

$$\leq (E[L^2])^{1/2}\|u^\nu - u\|\|Q^\nu\|_2 + \left| \int g(u,\omega)[Q^\nu(\omega) - Q(\omega)]d\mathbb{P}(\omega) \right|.$$

By the Uniform Boundedness Principle, $\{\|Q^\nu\|_2\}_{\nu=1}^\infty$ is bounded and the first term therefore vanishes. Since assumptions (i) and (iii) imply that $g(u,\omega) \in \mathcal{L}^2$ for all $u \in U$, the second term vanishes by the weak convergence of Q^ν to Q.

Second we consider $\nabla_u h$. Following a standard argument and the Dominated Convergence Theorem (see for example the proof of Theorem 7.44 in [33]), we find that for every $u \in U$ and $Q \in \mathcal{L}^2$, $\nabla_u h(u, Q)$ exists and is given by

$$\nabla_u h(u, Q) = \int \nabla_u g(u, \omega)Q(\omega)d\mathbb{P}(\omega).$$

Repeating the above argument with g replaced by $\partial g/\partial u_i$ and assumption (i) by assumption (ii) establishes the claim about $\nabla_u h$. $\square$

In view of Proposition 4.1, the following conclusions is a direct consequence of [29, Theorem 10.31].

4.2 Theorem (subdifferentiability of f). *Suppose that the assumptions of Proposition 4.1 holds. Then, f in (18) is locally Lipschitz continuous on U and strictly differentiable[11] where it is differentiable. There exists a set $D \subset U$ such that $U \setminus D$ is negligible[12], f is differentiable on D, and the gradient ∇f is continuous relative to the set D.*

Moreover, the directional derivative of f at $u \in U$ in direction $v \in \mathbb{R}^n$ is

$$df(u)(v) = \max \left\{ \langle E\left[\nabla_u g(u,\cdot)Q\right], v \rangle \ \middle| \ Q \in \mathcal{Q}^{g(u,\cdot)} \right\}$$

---

[11]Recall that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is strictly differentiable at a point $\bar{x}$ if $f(\bar{x})$ is finite and there is a vector $v \in \mathbb{R}^n$ such that $(f(x') - f(x) - \langle v, x' - x \rangle)/|x' - x| \to 0$ whenever $x, x' \to \bar{x}$ and $x' \neq x$; see [29, Definition 9.17].
[12]A subset of a set of Lebesgue measure zero is negligible.

and the subdifferential of f at $u \in U$ is

$$\partial f(u) = \text{con} \left\{ E\left[\nabla_u g(u, \cdot) Q\right] \;\middle|\; Q \in \mathcal{Q}^{g(u, \cdot)} \right\},$$

where $\mathcal{Q}^{g(u, \cdot)}$ is given in (16) with X replaced by $g(u, \cdot)$.

$\square$

We observe that when $\lambda = \bar{\lambda}_\alpha$, i.e., the focus is on a second-order superquantile risk measure $\bar{\mathcal{R}}_\alpha$, then $\mathcal{Q}^{g(u, \cdot)}$ is fully characterized by Theorem 3.8. In particular, the latter half of that theorem provides a specific risk identifier $Q \in \mathcal{Q}^{g(u, \cdot)}$ that is easily calculated when Ω has finite cardinality. Such a risk identifier then provides the subgradient $E[\nabla_u g(u, \cdot) Q]$ of f, which also is easily calculated in this case.

4.2 Application to Superquantile Regression

Superquantile regression as laid out in [23] (see also [21] and [14], which also includes several applications) resembles quantile regression, but instead of estimating conditional quantiles focuses on conditional superquantiles; quantifies of substantial interest as demonstrated by the numerous references listed in [23]. In particular, the approach of estimating a weighted average of conditional quantiles (see for example [5]) can be viewed as an approximate approach to estimating conditional superquantiles as realized from (1). The approach given in [23] is more direct and avoid approximations of this kind. Specifically, we find that for $Y \in \mathcal{L}^2$ and $\alpha \in (0, 1)$,

$$\{\bar{q}_Y(\alpha)\} = \operatorname*{argmin}_{u_0 \in \mathbf{R}} \bar{\mathcal{E}}_\alpha(Y - u_0), \text{ where } \bar{\mathcal{E}}_\alpha(Y) := \bar{\mathcal{V}}_\alpha(Y) - E[Y]$$

is a *measure of error* given in terms of the *measure of regret* [13] $\bar{\mathcal{V}}_\alpha$ defined in (5). In the same manner as minimizing mean-squared error yields an expectation and the foundation for least-squares regression, and minimizing a Koenker-Basset error yields a quantile and the foundation for quantile regression, minimizing $\bar{\mathcal{E}}_\alpha$ leads to superquantile regression.

Superquantile regression deals with the problem of approximating a random variable $Y \in \mathcal{L}^2$ by a combination of more accessible random variables $X_1, X_2, ..., X_n \in \mathcal{L}^2$, such that the error as quantified by $\bar{\mathcal{E}}_\alpha$ is minimized. Hopefully, the knowledge of $X = (X_1, ..., X_n)$ would then provide reasonably accurate predictions of Y. We stress that the terminology "regression" is here used more broadly than in the statistics literature. Limiting the scope to affine regression functions, superquantile regression then needs to solve the problem

$$\min_{u_0 \in \mathbf{R}, u \in \mathbf{R}^n} \quad \bar{\mathcal{E}}_\alpha \left(Y - [u_0 + \langle u, X \rangle] \right)$$

to obtain regression coefficients u_0 and u. That is, the regression coefficients (u_0, u) are selected such that the error between Y and the model $u_0 + \langle u, X \rangle$ is minimized.

We show in [23] that this problem can be decomposed into the two problems

$$\text{(i) find } \hat{u} \in \operatorname*{argmin}_{u \in \mathbf{R}^n} \frac{1}{1 - \alpha} \int_\alpha^1 \bar{q}_{g(u, \cdot)}(\beta) d\beta - E[g(u, \cdot)] \text{ and (ii) find } \hat{u}_0 = \bar{q}_{g(\hat{u}, \cdot)}(\alpha),$$

---

[13] We refer to [26] for a general treatment of measures of error and regret.

where for each $u \in I\!R^n$,

$$g(u, \cdot) = Y - \langle u, X \rangle$$

is a random variable defined on the sample space $\Omega = I\!R^{n+1}$, with sigma-algebra $\mathcal{B}_{R^{n+1}}$, and probability $\mathbb{P}$ given by the distribution of (X, Y). The problem (i) is that of minimizing a second-order superquantile of $g(u, \cdot)$ minus the expectation of $g(u, \cdot)$. Since $E[g(u, \cdot)] = E[Y] - \langle u, E[X] \rangle$ is a deterministic quantity, this problem is essentially in the form discussed earlier in the section: to minimize a mixed superquantile risk measure, in fact a second-order superquantile risk measure.

Suppose that the distribution $\mathbb{P}$ is supported on the points $\{(x^j, y^j)\}_{j=1}^{\nu} \subset I\!R^{n+1}$ with $\mathbb{P}\{(x^j, y^j)\} = p^j$, $j = 1, ..., \nu$, as is the case in practice when the regression relies on the observed data $\{(x^j, y^j)\}_{j=1}^{\nu}$. Then, the evaluation at a given $u \in I\!R^n$ of the objective function

$$f(u) = \frac{1}{1-\alpha} \int_\alpha^1 \bar{q}_{g(u,\cdot)}(\beta) d\beta - E[g(u, \cdot)]$$

of problem (i) and a corresponding subgradient are achieved as follows: Determine the cumulative distribution function of $g(u, \cdot)$ and use the formula in the second half of Theorem 3.8, with X replaced by $g(u, \cdot)$, to determine a risk identifier $\bar{Q}_\alpha^{g(u,\cdot)}$. This computation can be obtained in $O(\nu \log \nu)$ time, with sorting of $\{y^j - \langle u, x^j \rangle\}_{j=1}^{\nu}$ to obtain the cumulative distribution function being the bottleneck. Then, in view of Theorem 4.2, the function value and a subgradient are readily available through

$$f(u) = \sum_{j=1}^{\nu} p^j (y^j - \langle u, x^j \rangle) \bar{Q}_\alpha^{g(u,\cdot)}(\omega^j) - \sum_{j=1}^{\nu} p^j (y^j - \langle u, x^j \rangle)$$

and

$$\nabla f(u) = \sum_{j=1}^{\nu} -p^j x^j \bar{Q}_\alpha^{g(u,\cdot)}(\omega^j) + \sum_{j=1}^{\nu} p^j x^j, \text{ where } \omega^j = (x^j, y^j).$$

We note that the assumptions of Proposition 4.1 are easily verified in this case due, in part, to the affine form of $g(\cdot, \omega)$. Consequently, each iteration of a cutting-plane method or bundle method requires therefore computational time of order $O(\nu \log \nu)$ as a function of the number of data points. The number of iterations needed would depend on the method, n (the number of explanatory variables), and other factors. In comparison, a "primal" method proposed in [23] for solving the same problem requires the solution of a linear program with $n + O(\nu^2)$ variables and $O(\nu^2)$ inequality constraints. It is therefore clear that for small n and large ν, which is typical in regression problems, a dual method relying on the expressions derived in this paper might outperform the linear-programming-based approach; see [14] for empirical evidence supporting this claim. In fact, even storage of the linear program becomes challenging for large ν.

A   Appendix

As support for proving Proposition 2.3 in Section 2, we need the following consequence of the Fubini-Tonelli's Theorem.

A.1 Proposition *Suppose that $(\mathcal{X}, \mathcal{A}, \mu)$ and $(\mathcal{Y}, \mathcal{B}, \nu)$ are sigma-finite measure spaces. If $f : \mathcal{X} \times \mathcal{Y} \to \overline{I\!R}$ is measurable with respect to the product sigma-algebra on $\mathcal{X} \times \mathcal{Y}$ and $g : \mathcal{X} \times \mathcal{Y} \to \overline{I\!R}$ is integrable*

with respect to the product measure $\mu \times \nu$, with $f(x, y) \geq g(x, y)$ for $(\mu \times \nu)$-a.e. $(x, y) \in \mathcal{X} \times \mathcal{Y}$, then the following hold:

(i) the function $h_1 = \int f(x, \cdot) \, d\mu(x)$ is $\mathcal{B}$-measurable,

(ii) the function $h_2 = \int f(\cdot, y) \, d\nu(y)$ is $\mathcal{A}$-measurable,

(iii) and

$$\int f \, d(\mu \times \nu) = \int \left[\int f(x, y) \, d\mu(x) \right] d\nu(y) = \int \left[\int f(x, y) \, d\nu(y) \right] d\mu(x).$$

Proof. We recall that the integral of the sum of a nonnegative measurable function and an integrable function equates the sum of the individual integrals under the usual rules for handling addition with infinity. Then,

$$h_1 = \int f(x, \cdot) d\mu(x) = \int (f - g)(x, \cdot) d\mu(x) + \int g(x, \cdot) d\mu(x)$$

is $\mathcal{B}$-measurable since both terms on the right-hand side are $\mathcal{B}$-measurable by the Fubini-Tonelli Theorem. A similar argument yields the conclusion for h_2. The final assertion follows by applying the Fubini-Tonelli Theorem to $f - g$ and g, and the above rule about interchange of summation and integration. $\square$

Proof of Proposition 2.3. For every $X \in \mathcal{L}^2$, $\bar{q}_X$ is continuous and finite on $[0, 1)$ and therefore $\bar{\mathcal{B}}_{[0,1)}$-measurable. Moreover, $\bar{q}_X \geq E[X]$ and therefore $\mathcal{R}(X) \geq E[X] > -\infty$. Consequently, $\mathcal{R}$ is well-defined with values in $[E[X], \infty]$. Its regularity and positive homogeneity follow directly from those of $\mathcal{R}_\alpha$; see [26]. Since $\bar{q}_X$ is strictly increasing on $[0, 1)$ for nonconstant X, we have that if $\lambda(\{0\}) < 1$, then

$$\mathcal{R}(X) = E[X]\lambda(\{0\}) + \int_{1 > \beta > 0} \bar{q}_X(\beta) \, d\lambda(\beta) > E[X]\lambda(\{0\}) + E[X](1 - \lambda(\{0\})) = E[X]$$

and the strict lower bound follows. From (6),

$$\mathcal{R}(X) \leq \int_0^1 E[X] + \frac{\sigma(X)}{\sqrt{1 - \beta}} \, d\lambda(\beta) = E[X] + \sigma(X) \int_0^1 \frac{1}{\sqrt{1 - \beta}} \, d\lambda(\beta) < \infty$$

under the stated assumption, which establishes the corresponding finiteness on $\mathcal{L}^2$. In the case of $\sup X < \infty$, finiteness of $\mathcal{R}(X)$ follows trivially.

We next consider the alternative expression. By definition,

$$\mathcal{R}(X) = \int_0^1 \left[\int_0^1 q_X(\beta)\psi(\alpha, \beta) d\beta \right] d\lambda(\alpha), \tag{20}$$

with $\psi(\alpha, \beta) = \frac{1}{1-\alpha}$ if $0 \leq \alpha < \beta < 1$ and $\psi(\alpha, \beta) = 0$ otherwise. We equip $[0, 1) \times (0, 1)$ with the product measure $\lambda \times m$ defined on the product sigma-algebra $\bar{\mathcal{B}}_{[0,1)} \otimes \mathcal{B}_{(0,1)}$. It is obvious that $\psi : [0, 1) \times (0, 1) \to \mathbb{R}$ is $(\bar{\mathcal{B}}_{[0,1)} \otimes \mathcal{B}_{(0,1)})$-measurable and likewise q_X, viewed as a function on $[0, 1) \times (0, 1)$ that is constant in its first argument, due its monotonicity. Consequently, the function $(\alpha, \beta) \mapsto q_X(\beta)\psi(\alpha, \beta)$ is measurable in the same sense. Then, we look toward the interchange of integration order in (20).

22

We consider three cases. (i) Suppose that $X \geq 0$ a.e. Then, $q_X \geq 0$ and $q_X \psi \geq 0$, and the interchange of integration order is permitted by Tonelli-Fubini's Theorem. (ii) Suppose that $X \leq 0$ a.e. Then, $-q_X \geq 0$ and $-q_X \psi \geq 0$, and the interchange of integration order is again permitted by Tonelli-Fubini's Theorem. (iii) Suppose that neither (i) nor (ii) holds. Then, there exists a $\beta_X \in (0,1)$ such that $q_X(\beta) \geq 0$ for $\beta \geq \beta_X$ and $q_X(\beta) \leq 0$ for $\beta \leq \beta_X$. In view of Proposition A.1, it suffices to find an integrable, lower-bounding function of $q_X \psi$. Let $g : [0,1) \times (0,1) \to I\!R$ be given by

$$
g(\alpha, \beta) = \begin{cases} q_X(\beta)/(1 - \beta_X) & \text{if } 0 \leq \alpha < \beta \leq \beta_X \\ q_X(\beta) & \text{if } 0 \leq \alpha < \beta < 1, \beta_X < \beta \\ 0 & \text{otherwise.} \end{cases}
$$

Clearly, $q_X \psi \geq g$ and

$$
\int |g| d(\lambda \times m) \leq \frac{1}{1 - \beta_X} \int |q_X| d(\lambda \times m) = \frac{1}{1 - \beta_X} \int_0^1 \left[\int_0^1 |q_X(\beta)| d\beta \right] d\lambda(\alpha), \tag{21}
$$

where the equality follows by Tonelli-Fubini's Theorem. The inner integral simplifies to

$$
\int_0^1 |q_X(\beta)| d\beta = \int_{\beta_X}^1 q_X(\beta) d\beta - \int_0^{\beta_X} q_X(\beta) d\beta = (1 - \beta_X)\bar{q}_X(\beta_X) - \int_0^{\beta_X} q_X(\beta) d\beta.
$$

The last term requires further simplification. Recall that for $\alpha \in (0,1)$,

$$
\frac{1}{\alpha} \int_0^\alpha q_X(\beta) d\beta = -\frac{1}{\alpha} \int_{1-\alpha}^1 q_{-X}(\beta) d\beta = -\bar{q}_{-X}(1 - \alpha).
$$

Applying this result, the inner integral from above simplifies further to

$$
\int_0^1 |q_X(\beta)| d\beta = (1 - \beta_X)\bar{q}_X(\beta_X) + \beta_X \bar{q}_{-X}(1 - \beta_X) < \infty.
$$

Consequently in view of (21), g is integrable and therefore furnishes the necessary lower-bounding, integrable function in Proposition A.1, which completes part (iii). We are therefore permitted to interchange the order of integration in (20) and get

$$
\mathcal{R}(X) = \int_0^1 \left[\int_0^1 q_X(\beta) \psi(\alpha, \beta) d\beta \right] d\lambda(\alpha) = \int_0^1 q_X(\beta) \left[\int_0^1 \psi(\alpha, \beta) \, d\lambda(\alpha) \right] d\beta = \int_0^1 q_X(\beta) \varphi(\beta) d\beta,
$$

where the last equality follows from the definition of φ.

The final assertions follow from recognizing that the Lebesgue-Stieltjes measure $d\varphi$ associated with a function φ has $d\varphi(\alpha) = \frac{1}{1-\alpha} \, d\lambda(\alpha)$ for a weighting measure λ on $[0,1)$. $\qquad\square$

Now we articulate other definitions and technical results required in the paper.

A.2 Definition *Let $(T, \mathcal{A}, \mu)$ be a complete measure space, with μ sigma-finite, $\mathcal{X}$ a separable reflexive Banach space, and $\mathcal{M}$ a linear subspace of the linear space of all $(\mathcal{A}, \mathcal{B}_\mathcal{X})$-measurable functions $x : T \to$*

$\mathcal{X}$. The set $\mathcal{M}$ is $(\mathcal{A}, \mathcal{B}_{\mathcal{X}})$-decomposable if, whenever $x \in \mathcal{M}$ and $x_0 : S \to \mathcal{X}$ is a bounded $(\mathcal{A}, \mathcal{B}_{\mathcal{X}})$-measurable function on a set $S \in \mathcal{A}$, with $\mu(S) < \infty$, then the function $y : T \to \mathcal{X}$ given by

$$y(t) = \begin{cases} x_0(t) & \text{if } t \in S \\ x(t) & \text{if } t \in T \setminus S \end{cases}$$

also belongs to $\mathcal{M}$.

A.3 Definition *In the notation of Definition A.2, we say that a function $f : T \times \mathcal{X} \to (-\infty, \infty]$ is a* normal integrand *if the following hold:*

(i) f *is* $(\mathcal{A} \otimes \mathcal{B}_{\mathcal{X}})$-measurable and

(ii) *for every* $t \in T$, $f(t, \cdot)$ *is lower semicontinuous on $\mathcal{X}$ and not identical to ∞.*

A.4 Proposition *Suppose that the conditions and notation of Definition A.2 hold and $f : T \times \mathcal{X} \to (-\infty, \infty]$ is a normal integrand. Then, the following hold:*

(i) *the functions $t \mapsto \inf_{\xi \in \mathcal{X}} f(t, \xi)$ and $t \mapsto f(t, x(t))$, with $x : T \to \mathcal{X}$ $(\mathcal{A}, \mathcal{B}_{\mathcal{X}})$-measurable, are $\mathcal{A}$-measurable and*

(ii) *if $\mathcal{M}$ is $(\mathcal{A}, \mathcal{B}_{\mathcal{X}})$-decomposable and there exists an $x \in \mathcal{M}$ such that $\int f(t, x(t)) \, d\mu(t) < \infty$, then*

$$\inf_{x \in \mathcal{M}} \int f(t, x(t)) \, d\mu(t) = \int \varphi(t) \, d\mu(t), \text{ where } \varphi(t) = \inf_{\xi \in \mathcal{X}} f(t, \xi). \tag{22}$$

Proof. First, we consider $t \mapsto \inf_{\xi \in \mathcal{X}} f(t, \xi)$. For measurable spaces $(\mathcal{X}_1, \mathcal{A}_1)$ and $(\mathcal{X}_2, \mathcal{A}_2)$, we recall that a set-valued mapping $S : \mathcal{X}_1 \rightrightarrows \mathcal{X}_2$ is $(\mathcal{A}_1, \mathcal{A}_2)$-measurable if its graph is measurable in the sense that

$$\{(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \mid x_2 \in S(x_1)\} \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$

where $\mathcal{A}_1 \otimes \mathcal{A}_2$ is the product sigma-algebra generated by $\mathcal{A}_1$ and $\mathcal{A}_2$. Since f is a normal integrand, the set-valued mapping $t \mapsto \text{epi } f(t, \cdot)$ is $\mathcal{A}$-measurable and closed-valued; see for example [18, Proposition 1]. By [18, Theorem 1(f)], there exists a countable collection $\{g_i\}_{i \in I}$ of $\mathcal{A}$-measurable functions $g_i : T \to \mathcal{X} \times \mathbb{R}$ of the form $g_i(t) = (x_i(t), \alpha_i(t))$, $x_i(t) \in \mathcal{X}$ and $\alpha_i(t) \in \mathbb{R}$, such that

$$\text{epi } f(t, \cdot) = \text{cl}\{g_i(t)\}_{i \in I} \text{ for all } t \in T,$$

where cl denotes closure. The mapping $t \mapsto \alpha_i(t)$ is also $\mathcal{A}$-measurable. Consequently,

$$\inf_{\xi \in \mathcal{X}} f(t, \xi) = \inf_{i \in I} \alpha_i(t) \text{ for all } t \in T$$

and the conclusion follows from the fact that the pointwise infimum of a countable collection of measurable functions is a measurable function.

Second, we consider $t \mapsto f(t, x(t))$, which is a composition of f with the measurable mapping $t \mapsto (t, x(t))$ and therefore measurable.

24

Third, we establish part (ii) by following the arguments in the proof of Theorem 2 in [18]. By assumption there exists a function $x_1 \in \mathcal{M}$ and a μ-integrable function $\alpha_1 : T \to I\!\!R$ such that

$$f(t, x_1(t)) \leq \alpha_1(t) \text{ for every } t \in T.$$

Since $\varphi(t) \leq f(t, x(t))$ for every function $x \in \mathcal{M}$ and $t \in T$ by definition and φ is $\mathcal{A}$-measurable by part (i), the integral of φ is well-defined and either finite or equals $-\infty$. Consequently, the inequality $\geq$ holds in (22). Now, let $\gamma \in I\!\!R$ be such that

$$\int \varphi(t) \, d\mu(t) < \gamma. \tag{23}$$

We will prove the existence of a function $x \in \mathcal{M}$ such that

$$\int f(t, x(t)) \, d\mu(t) < \gamma, \tag{24}$$

thereby establishing part (ii). From (23) and the properties of $(T, \mathcal{A}, \mu)$, there exists a μ-integrable function $\alpha_0 : T \to I\!\!R$ such that $\varphi(t) < \alpha_0(t)$ for every $t \in T$ and

$$\int \alpha_0(t) \, d\mu(t) < \gamma. \tag{25}$$

We define the set-valued mapping $S : T \rightrightarrows \mathcal{X}$ by

$$S(t) = \{\xi \in \mathcal{X} \mid f(t, \xi) \leq \alpha_0(t)\} \quad \text{for } t \in T.$$

Since the function $(t, \xi) \mapsto f(t, \xi) - \alpha_0(t)$ is $(\mathcal{A} \otimes \mathcal{B}_{\mathcal{X}})$-measurable, S is also $\mathcal{A}$-measurable. Moreover, $S(t)$ is for each $t \in T$ closed and nonempty. Since S is $\mathcal{A}$-measurable, there exists a $\mathcal{A}$-measurable selection x_0, i.e., a $\mathcal{A}$-measurable function x_0 such that $x_0(t) \in S(t)$ for every $t \in T$; see for example the corollary of Theorem 1 in [18]. Since (25) holds, there exists a measurable set $T_0 \subset T$, with $\mu(T_0) < \infty$, such that

$$\int_{T_0} \alpha_0(t) \, d\mu(t) + \int_{T \setminus T_0} \alpha_1(t) \, d\mu(t) < \gamma. \tag{26}$$

By the construction of S in terms of α_0, the measurable selection x_0 can be chosen to be bounded on T_0. Let $x : T \to \mathcal{X}$ be such that $x(t) = x_0(t)$ for $t \in T_0$ and $x(t) = x_1(t)$ for $t \in T \setminus T_0$. Then, $x \in \mathcal{M}$ by the assumption of decomposability, and we have that $f(t, x(t)) \leq \alpha_0(t)$ for $t \in T_0$ and $f(t, x(t)) \leq \alpha_1(t)$ for $t \in T \setminus T_0$. From (26) we then conclude (24), which establishes part (ii). $\qquad \square$

A.5 Lemma *If $q : [0, 1) \to \mathcal{L}^2$ is $(\bar{\mathcal{B}}_{[0,1)}, \mathcal{B}_{\mathcal{L}^2})$-measurable, then*

  (i) *the function $f_1 : [0, 1) \times \Omega \to \overline{I\!\!R}$ given by $f_1(\beta, \omega) = q(\beta)(\omega)$ is $(\bar{\mathcal{B}}_{[0,1)} \otimes \mathcal{F})$-measurable, and*

  (ii) *the function $f_2 : [0, 1) \to I\!\!R$ given by $f_2(\beta) = \|q(\beta)\|_2$ is $\bar{\mathcal{B}}_{[0,1)}$-measurable.*

Proof. For part (i) simply observe that $f_1 = g \circ h$, where $h : [0, 1) \times \Omega \to \mathcal{L}^2 \times \Omega$, with $h(\alpha, \omega) = (q(\alpha), \omega)$, and $g : \mathcal{L}^2 \times \Omega \to \overline{I\!\!R}$, with $g(Q, \omega) = Q(\omega)$. The conclusion then follows from the measurability of q and elements of $\mathcal{L}^2$, and the fact that composition of measurable functions is measurable. Next,

we consider part (ii). A trivial extension of part (i) establishes that the function $(\beta, \omega) \mapsto [q(\beta)(\omega)]^2$ is $(\bar{\mathcal{B}}_{[0,1)} \otimes \mathcal{F})$-measurable. Since it is also nonnegative, it follows from Tonelli-Fubini's Theorem that $[f_2(\cdot)]^2$ is $\bar{\mathcal{B}}_{[0,1)}$-measurable. $\square$

The following is a direct consequence of Definition A.2.

A.6 Proposition *The set $\mathcal{M}$ is $(\bar{\mathcal{B}}_{[0,1)}, \mathcal{B}_{\mathcal{L}^2})$-decomposable.*

References

[1] C. Acerbi. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26:1505–1518, 2002.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[3] C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions*. Springer, 1977.

[4] P. Cheridito and T. Li. Dual characterization of properties of risk measures on orlicz hearts. *Mathematics and Financial Economics*, 2(1):29–55, 2008.

[5] S. Y. Chun, A. Shapiro, and S. Uryasev. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research*, 60(4):739–756, 2012.

[6] R. Datko. On the integration of set-valued mappings in a Banach space. *Fundamenta Mathematicae*, 78:205–208, 1973.

[7] H. Föllmer and A. Schied. *Stochastic Finance*. de Gruyter, 1. edition, 2002.

[8] H. Föllmer and A. Schied. *Stochastic Finance*. de Gruyter, 2. edition, 2004.

[9] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. de Gruyter, 3. edition, 2011.

[10] E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the Fatou property. In S. Kusuoka and T. Maruyama, editors, *Advances in Mathematical Economics, Volume 9*, pages 49–71. Springer, 2006.

[11] S. Kusuoka. On law-invariant coherent risk measures. In S. Kusuoka and T. Maruyama, editors, *Advances in Mathematical Economics, Volume 3*, pages 83–95. Springer, 2001.

[12] H. Mausser and D. Rosen. Efficient risk/return frontiers for credit risk. *Algo. Research Quarterly*, 2(4):35–47, 1998.

[13] N. Miller and A. Ruszczynski. Risk-averse two-stage stochastic linear programming: modeling and decomposition. *Operations Research*, 59(1):125–132, 2011.

[14] S. Miranda. *Superquantile regression: theory, algorithms, and applications*. PhD thesis, Naval Postgraduate School, Monterey, CA, 2014.

[15] N. Noyan and G. Rudolf. Kusuoka representations of coherent risk measures in general probability spaces. *Optimization Online*, 2013.

[16] G. Ch. Pflug. On distortion functionals. *Statistics and Decisions*, 24:45–60, 2006.

[17] G. Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk.* World Scientific Publishing, 2007.

[18] R. T. Rockafellar. Convex integral functionals and duality. In E. Zaratonello, editor, *Contributions to Nonlinear Functional Analysis*, pages 215–236. Academic Press, 1971.

[19] R. T. Rockafellar and J. O. Royset. On buffered failure probability in design and optimization of structures. *Reliability Engineering & System Safety*, 95:499–510, 2010.

[20] R. T. Rockafellar and J. O. Royset. Superquantiles and their applications to risk, random variables, and regression. *INFORMS, Tutorials in Operations Research*, pages 151–167, 2013.

[21] R. T. Rockafellar and J. O. Royset. Random variables, monotone relations and convex analysis. *Mathematical Programming B*, 148:297–331, 2014.

[22] R. T. Rockafellar and J. O. Royset. Measures of residual risk with connections to regression, risk tracking, surrogate models, and ambiguity. *SIAM J. Optimization*, 25(2):1179–1208, 2015.

[23] R. T. Rockafellar, J. O. Royset, and S.I. Miranda. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research*, 234(1):140–154, 2014.

[24] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[25] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[26] R. T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18:33–53, 2013.

[27] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Deviation measures in risk analysis and optimization. Technical report, Department of Industrial and Systems Engineering, University of Florida, 2002.

[28] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10:51–74, 2006.

[29] R. T. Rockafellar and R. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaft.* Springer, 3rd printing-2009 edition, 1998.

[30] A. Ruszczynski. Advances in risk-averse optimization. In H. Topaloglu, editor, *Tutorials in Operations Research*, pages 168–190. INFORMS, 2013.

[31] A. Ruszczynski and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.

[32] A. Shapiro. On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, 38(1):142–152, 2013.

[33] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.